(* TWO-WAY *)

20

Blocks and unification
(organized comparisons)

	Exh	Page
Introduction 20-1		
block, at 20-2		
	A	
	1	20-8
	B	
20A. Three things compared in pairs -- direct and indirect comparisons 20-5	2	20-11
	3	20-13/14
	C	
20B. Combination of comparisons 20-9	4	20-16
- centering 20-10	5	20-19
combining residuals, at 20-12	6	20-21
	D	
20C. Complete comparisons 20-15	7	20-23
	8	20-24
20D. The full carbon-paper experiment 20-22	9	20-27
- polished combination 20-25	10	20-29
- comparing the combining residuals 20-28	11	20-30
	12	20-31
20E. More than two, but less than all 20-32		
	E	
	13	20-33/34
20F. Individual residuals and two-way analyses 20-37	14	20-36
- more general uses 20-45		
	F	
- transparency and insulation 20-49	15	20-39/40
	16	20-41
20G. "Weighting" by polishing 20-49	17	20-42/43
	18	20-44
**********	19	20-47/48
	G	
20H. How far have we come?	20	20-50
	21	20-52/53
**********		
	H	
	K	
20K. Where the unifying table has to wait (optional) 20-57	22	20-58/61
	23	20-62
**********		
	P	
	24	20-65
20P. Problems for chapter 20 20-63	25	20-66
	26	20-67
	27	20-68
	28	20-69
	29	20-70
	30	20-71
	31	20-72
	32	20-73

20

Blocks and unification (organized comparisons)

The Greeks probably had a word for it; the Romans used ceteris paribus -- "other things being equal" -- thus recognizing the chief essential of effective comparison, to compare A with B with other things at least similar, not with other things varying unnecessarily.

It is rare indeed that only the differences that we are studying affect the response through whose values we are to make the comparison. It is equally rare that we know all the other things that do matter. We want as clear and precise comparisons as we can get. The price of this is "holding fixed" those things which:

- we can reasonably control.
- and seem likely to affect the response.

For a number of reasons, we are likely to hold these other things fixed in different ways, comparing A with B with other things at γ, and with other things at δ, and with other things at ε, and so on. Two of the important reasons for this are:

- it is often hard to hold other things constant, either by finding or by altering, over enough time, or over enough space, or over enough people, or over enough experimental plants or animals, to allow us

to make enough comparisons to give us enough precision in our results.

- holding other things constant in only one way tends to make our results more special in their application; holding other things constant in each of many diverse ways tends to make our results much more general in their application.

Because careful observation and experimentation of this sort was first written about seriously in agriculture, the most used term for the group of situations in which other things are held constant is a

block.

When growing plants, a farmer uses land. Land varies in its characteristics in important ways -- many of which we still do not know how to measure by anything simpler than growing plants. A major part of "holding other things constant" in an agricultural experiment is holding land characteristics constant.

The best way we know how to do this is still as was many years ago -- seek out some reasonably uniform land, and select a compact piece to be called a block. Then divide this block into plots. If we then try things to be compared on plots, and make our comparisons between plots that are:

- in the same block
- treated differently

we will have done a lot to make our comparisons more precise by reducing the impact of differences from one patch of land to another.

In chemical engineering experiments, by contrast, the unit is usually called a "run". Making a run uses a piece of equipment -- or several pieces in sequence -- for a certain section of time and requires an operator to try to maintain specified temperatures, pressures, and other meter readings. Here a block consists of a number of runs chosen to be as similar as we can easily attain: ordinarily using the same equipment at nearby times under the guidance of the same operator.

Sometimes keeping conditions reasonably constant requires some separation in time rather than working one right after another. It would probably be a mistake to block an experiment on mowing home lawns by days, for example. Lawns are different early in the morning when wet with dew than they are in midafternoon when well dried by the sun.

As reported in 1878?, the Great Trigonometric Survey of India gave up trying to carry levelling across the hot plains of India by sighting from two ends of a line segment at the same time. Instead they made their level sights near 1:00 o'clock in the afternoon when the heat of the day was well established. In effect, they used just one block "near 1 p.m.".

Organized experimentation or observation almost always gains from blocks, blocks that are rather easily identified as such. There may or may not be anything corresponding to plots or runs in the senses of agricultural and chemical-engineering experimentation. Observational studies of people, for instance, where people are first paired with respect to background characteristics such as age, sex, ethnic group, intelligence, etc. and where the differences among pairs in the factor to be studied are related to differences in outcome, which may be school performance, life satisfaction, or health, are an example. Each pair of persons is a block, but it is often not at all exact to think of each person as a plot or run. We usually have no choice as to which person in a pair "gets" which version of the factor we are studying. We still make comparisons within blocks, but we are exposed to additional problems of interpretation, since some factor we have not observed may really be the cause of the outcomes.

Careful study of most phenomena leads to some

organized comparisons

to situations where numbers are obtained -- and can be compared -- under different background circumstances. Almost always, comparisons between the versions of a factor that interest us are made more precisely and reliably within a set of circumstances -- within a block -- than they would be made from one block to another.

Once we have made comparisons by pairs of versions there is a desire to give numbers to each version separately, numbers whose differences do well in comparing the corresponding pairs of versions. A problem arises when not all versions are compared in every block. A similar one often comes up when all versions do not appear in every block.

The techniques we learn in handling blocked data will turn out to be relevant to the analysis of other two-way tables, especially those that are highly incomplete because many values are absent.

20A. Three things compared in pairs -- direct and indirect comparisons

In 1958, Fleckenstein, Freund, and Jackson reported an experiment to compare brands of carbon paper. Their approach was realistic: carbon paper is used in typing, mainly by secretaries and typists. It seemed reasonable to give an individual typist sample material to be typed with each of two brands of carbon paper, asking her to report her preference for one brand over the other on a seven-point scale from -3 to +3.

Giving an individual typist as many as three or more brands seemed likely to yield less precise judgments. If, on the other hand, each secretary had been given only one brand, differences from secretary to secretary in what specific words or a grade meant in individual reports could not be balanced out.

There are many other circumstances where comparison in pairs seems the best that can be done. Comparisons between insect repellants are an obvious example, since nothing is more alike than the two sides of the same cow or the two arms of the same man. Blocks of two are not uncommon.

The order in which a secretary typed the test material could influence her report. To balance any such effect, Fleckenstein, Freund, and Jackson arranged for the secretaries who compared two specific brands, say C and D, to be divided equally, one subgroup using C first, the other reversing the order -- just as major league baseball teams share home games.

We shall not go into as much detail about the results of this careful experiment as does the published paper. We shall focus on the issues that are central for us here. For each pair of brands, 30 secretaries gave preference scores on a numerical scale. Thus there were 30 preference scores that compared this pair of brands directly. Fleckenstein, Freund, and Jackson give the details of these preference scores, but it will be enough for us here to use a summary value for each pair compared. We could use their mean, but it is of course equivalent and easier to use their sum.

Depending on how we feel about the importance of secretaries with quite unusual opinions, we will have done better or worse to use sums or means of the batches of 30 preference scores rather than their trimeans or medians.

Exhibit 1 shows the results for brands C, D, and E. The first panel shows the direct comparisons. We need to remember that +61 for "C minus D" means -(+61) = -61 for "D minus C" and so on.

Suppose that the direct comparison of C with D had been lost. What would we use to replace it? We would still have comparisons of both C and D with E. If C was +13 as compared to E while D was -40, it seems reasonable to form the difference (+13) - (-40) = +53 and say that this, too, tells us something about C as compared to D. The second panel of exhibit 1 gives the three indirect comparisons that this data provides. In forming indirect comparisons, we often need a direct comparison turned around from the direct direction given in panel A. After all, we must follow the pattern "U with X" minus "V with X" tells us about "U with V".

exhibit 1 of chapter 20

Direct comparisons of three brands of carbon paper, and some steps toward analysis

A) DIRECT COMPARISONS -- sums of 30 ratings

Brand C minus Brand D: +61

Brand C minus Brand E: +13

Brand D minus Brand E: -40

B) INDIRECT COMPARISONS -- from direct comparison

C with D: (+13) - (-40) = +53

C with E: (+61) - (-(-40)) = +21

D with E: (-(+61)) - (-(+13)) = -48

C) COMBINED COMPARISONS -- counting direct comparisons double

C with D: $\frac{(+61) + (+61) + (+53)}{3} = 58.3$

C with E: $\frac{(+13) + (+13) + (+21)}{3} = 15.7$

D with E: $\frac{(-40) + (-40) + (-48)}{3} = -42.7$

D) SOURCE: Mary Fleckenstein, Richard A. Freund, and J. Edward Jackson 1958. A paired comparison test of typewriter carbon papers. Tappi 41: 128-130.

20B. Combination of comparisons

In this example, the indirect comparisons are in quite reasonable agreement with the direct comparisons. Clearly we should try to combine them. Equally clearly we trust the direct comparisons somewhat more. Accordingly, we will give each direct comparison twice the weight of an indirect comparison. The easy way to do this is to add direct comparisons in twice but indirect ones only once. This is done in panel C.

This particular way of combining direct and indirect comparisons comes with a bonus. The direct comparisons were not -- we did not expect them to be -- compatible in detail. We did not expect "C with E" of +13 and "E with D" of +40 to add up to exactly "C with D". Our combined comparisons, however, yield, for these three comparisons

$$15.7 + 42.7 = 58.4 \quad \text{versus} \quad 58.3$$

where the discrepancy clearly comes only from forcing thirds into tenths. This sort of combination guarantees compatibility, which is a bonus. We would not have had the bonus if we had weighted each direct comparison either more or less than twice as much as an indirect comparison.

Instead of our fit telling us separately about every comparison -- which can, in unusual circumstances, be necessary -- this bonus means that our fit tells us

directly about the individual versions, and that we are to calculate fits for comparisons by subtracting the fit for the second version from the fit for the first. This is little gain with 3 versions -- as in our example with 3 kinds of carbon paper -- since there are then only as many comparisons -- 3 -- as there are versions.

If there were 10 versions, however, there would be 45 comparisons, and it would be much easier to work with -- and think about -- fits involving only 10 numbers rather than with fits involving 45 separate numbers.

This reduction is something we always try for, and usually gain. In just the sort of experiment we have been discussing, however, there are situations where people sometimes act in a more complicated way, preferring C to D, D to E, and E to C, when given only pairs to choose among. We do not meet such situations often, but we need to keep their possibility in mind.

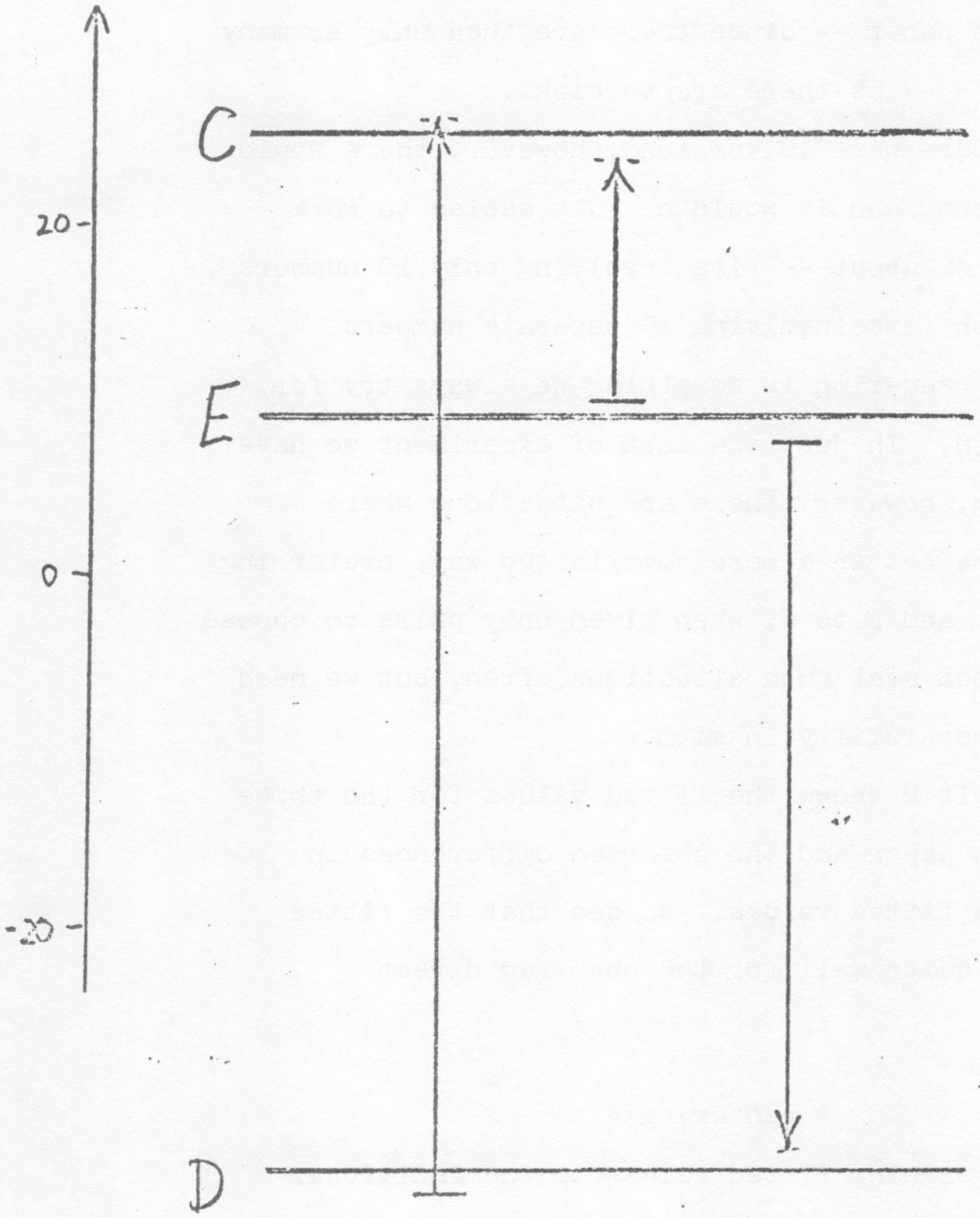
Exhibit 2 shows the fitted values for the three kinds of carbon paper and the observed differences in relation to the fitted values. We see that the fitted values account quite well for the observed direct comparisons.

* centering *

In assigning fitted values to the individual brands, we have one free choice. If we were to call C

exhibit 2 of chapter 20

Fitted values (horizontal lines) and
observed comparisons (arrows)
for 3 brands of carbon paper



"100" for example, then we would have to use 100 - 58.4 = 41.6 for D and 100 - 15.7 = 84.3 for E. Just how we make this choice does not matter. One convenient choice to make the "scores" for all candidates -- here C, D and E -- add to zero.

It may not be clear that this is a simple choice, but we shall see that it is.

If we are willing to live with this choice, we can make the calculation simply -- almost automatically -- by using

a square unifying table

as illustrated by exhibit 3. In this table we have laid out all nine comparisons of one of C, D, E with one of C, D, E. Clearly we are to put zero down for "C with C", "D with D" and "E with E". Equally, as we have done, we should use minus the result of "C with D" as the result of "D with C", and so on with the other possible cases of turning around.

Summing the columns in exhibit 3 and forming means, we get numbers whose differences are exactly the combined comparisons we saw in exhibit 1. These values were used to make the plot of exhibit 2. Their differences, shown in panel B can clearly be used to find the residuals shown in panel C. We specify that these residuals are

combining residuals

exhibit 3 of chapter 20

The "square unifying table" applied to the observed differences among carbon papers

Second brand	First brand C	First brand D	First brand E
A) The DIFFERENCES between FIRST BRAND and SECOND BRAND			
C	0	-61	-13
D	+61	0	+40
E	+13	-40	0
(sum)	(+74)	(-101)	(+27)
mean	+24.7	-33.7	+9.0
B) The FITTED DIFFERENCES			
C	0	-58.4	-15.7
D	+58.4	0	+42.7
E	+15.7	-42.7	0
(sum)	(+74.1)	(-101.1)	(+27.0)

Example: +58.4 = +24.7 - (-33.7)

C) COMBINING RESIDUALS, OBSERVED MINUS FITTED			
C	0	-2.6	+2.7
D	+2.6	0	-2.7
E	-2.7	+2.7	0
(sum)	(-0.1)	(+0.1)	(0.0)

Example: +2.6 = (+61) - (+58.4)

exhibit 3 (cont'd.)

D) CLARITY of COMBINATION -- omitting diagonal zeros

Range of fitted differences: 116.8

Midspread of combining residuals: 5.4

Clarity = 21^{+}

since they arise from combination and are not usually associated with individual observations. They are relatively small, in this example, since their midspread -- omitting the diagonal values that must vanish -- is only about 1/16 that for the fitted values.

The fact that the two residuals having Brand C as "first brand" add to zero (except for rounding) corresponds to the fact that, in exhibit 2, the observed differences associated with Brand C "balance" around the horizontal line for C. One difference would be better fitted if the line were higher, the other if it were lower. The structure of our analysis ensures this sort of balance. So long as we have only 3 things to compare, we have certainly done our best.

We call such a table a square unifying table because it is square and because -- by forcing the fitted comparisons to become compatible and unified -- it gains us the bonus of fitting individual versions rather than individual comparisons. It will prove to be a major tool.

20C. Complete comparisons

We do not always have to make comparisons two by two. Often we can compare all candidates together, repeating the whole comparison several times. How are we to look at such data?

Exhibit 4 sets out some of the results of an experiment on sugar beets reported by Snedecor (1946). The

exhibit 4 of chapter 20

Yields of sugar beets (tons/acre) in each of six blocks for three amounts of fertilizer

A) OBSERVED YIELDS

	Fertilizer applied		
	0, None	1, Phosphate only	2, Phosphate and Nitrate
Block 1	2.45	6.71	6.48
Block 2	2.25	5.44	7.11
Block 3	4.38	4.92	5.88
Block 4	4.35	5.23	7.54
Block 5	3.42	6.74	6.61
Block 6	3.27	4.74	8.86
Sum	20.12	33.78	42.48
Mean	3.35	5.63	7.08
(Median)	(3.34)	(5.34)	(6.86)

B) OBSERVED DIFFERENCES

	"1" minus "0"	"2" minus "1"	"2" minus "0"
Block 1	4.26	-0.23	4.03
Block 2	3.19	1.67	4.86
Block 3	0.54	0.96	1.50
Block 4	0.88	2.31	3.19
Block 5	3.32	-0.13	3.19
Block 6	1.47	4.12	5.59
Sum	13.66	8.70	22.36
Mean	2.28	1.45	3.73
(Median)	(2.33)	(1.32)	(3.61)

C) COMPARISONS -- means OR medians, differences before OR after

$$13.66 = 13.66 = 33.78 - 20.12$$
$$8.70 = 8.70 = 42.48 - 33.78$$
$$22.36 = 22.36 = 42.48 - 20.12$$

also

$$22.36 = 13.66 + 8.70$$

hence

$$2.28 = 2.28 = 5.63 - 3.35$$
$$1.45 = 1.45 = 7.08 - 5.63$$
$$3.73 = 3.73 = 7.08 - 3.35$$

but

$$2.33 \neq 2.00 = 5.34 - 3.34$$
$$1.32 \neq 1.52 = 6.86 - 5.34$$
$$3.61 \neq 3.52 = 6.86 - 3.34$$

D) SOURCE:

G. W. Snedecor 1946. Statistical Methods, 4th edition. Iowa State College Press, p. 332.

values given are yields in tons per acre for each of three fertilizer treatments applied to a plot in each of six blocks. Land varies. To make precise comparisons we should compare our fertilizer treatments on similar patches -- usually called plots -- of ground. To get similar patches, the best we can do in many situations is use patches that are close together. The simplest way to do this is to divide the field into blocks and then divide each block into a suitable number of plots. Our direct comparisons here are comparisons between plots within a single plot.

Panel A shows the yields, while panel B shows the direct comparisons. Sums and means are given for both. As panel C shows, we could find the mean difference between treatments by finding the mean response to each treatment and then taking differences. This works just so long as:

- we want to use means.
- the data is "balanced".

Clearly the next step in this analysis would be to form residuals. We know that we should be wary of residuals from means; we want to use medians or trimeans. The natural thing to do is to take medians or trimeans in either panel A or panel B. Which is better, and why?

Exhibit 5, involving hand-made numbers, clears up the main issue. Here the whole comparison has been repeated six times. Any set of values to be directly compared with one another -- in the way one directly compares plots in a block -- will be called a block whether or not it represents anything physically like a block of plots of land. Thus we have 6 blocks.

The left-hand side of exhibit 5 shows numbers which give a difference between treatments of +5 in EACH of the six blocks. No sensible analysis can give any overall comparison except 5. On the right-hand side, one "observed" number has been changed from +4 to -4. To be resistant, we ought not to respond very much to such a change of a single value. The result of this change is a change in the median for that treatment but no change in the median of the differences.

If we want our analysis to be resistant to such changes, we need to work with middle means of differences, not with differences of middle means.

If we are going to use medians in order to be resistant, then, panel B of exhibit 4 is the relevant panel.

exhibit 5 of chapter 20 -- hand-made

Hypothetical tables illustrating why we want medians of differences and not differences of medians

	Original		Modified	
A) "OBSERVED" VALUES				
	A	B	A	B
Block 1	10	5	10	5
Block 2	9	+4	9	-4
Block 3	5	0	5	0
Block 4	0	-5	0	-5
median	**7**	**2**	**7**	**-2**
B) "OBSERVED" DIFFERENCES "A MINUS B"				
Block 1	5		5	
Block 2	5		13	
Block 3	5		5	
Block 4	5		5	
median	**5**		**5**	
C) COMPARISON of VALUES of "A MINUS B" found from MEDIANS				
From panel A above	**5** = 7 - 2		**9** = 7 - (-2)	
From panel B above	**5**		**5**	

The indicated differences are

"1" minus "0": 2.33

"2" minus "1": 1.32

whose sum is 3.65, but

"2" minus "0": 3.61

Even though all comparisons were made in each block -- so that mean comparisons are compatible and unified -- the raw results of the median difference analysis do not give us the bonus of perfect compatibility. We know how to get this bonus, however. We need only use a square unifying table. Exhibit 6 shows the computations and the result. Clearly the fit finally obtained is again a good one -- this time the midspread of the combining residuals is only a very small fraction of that for the fitted differences.

If we want to represent the median comparisons of the various fertilizers as differences among values for treatments, it seems that the best we can do is such a choice as

0, No fertilizer added:	down 1.98
1, Phosphate only:	up 0.34
2, Phosphate and nitrate:	up 1.65

where these ups and downs are from a rather arbitrary zero.

exhibit 6 of chapter 20 -- sugar beets

The square unifying table applied to the median differences

Second treat.	First treatment: 0, None	1, Phosphate only	2, Phosphate and Nitrate
A) FIRST TREATMENT MINUS SECOND TREATMENT -- see panel B of exhibit 4			
0	0	2.33	3.61
1	-2.33	0	1.32
2	-3.61	-1.32	0
Sum	-5.94	1.01	4.93
Mean	-1.98	.34	1.64
B) FITTED DIFFERENCES -- found from fitted values			
0	0	2.32	3.62
1	-2.32	0	1.30
2	-3.62	-1.30	0
C) COMBINING RESIDUALS			
0	0	+.01	-.01
1	-.01	0	+.02
2	+.01	-.01	0

D) CLARITY of COMBINATION -- omitting diagonal zeros

Range of fitted differences 7.24

Midspread of combining residuals: 0.02

(Clarity: very large)

This arbitrariness is of little concern to us, so long as our concern is only with actual comparisons.

It is easy to be naive enough to feel that enough direct comparisons might eliminate any need for medians. That this is not so, even in the very simplest case where there is no need for even a unifying table, can be seen from an experiment of Mitchell, Burroughs, and Beadles on the biological value of proteins from peanuts, as shown by rat growth. Exhibit 7 sets out the data, with means and mid-means. Here, as a look at the ordered differences will convince almost everyone, the trimean difference of +3.8 -- or the median difference of 3.5 -- is by far a more useful summary of the results than the mean difference of +2.1, which is also the difference of means.

20D. <u>The full carbon-paper experiment</u>

As we might suspect from the notation C, D, E above, Fleckenstein, Freund, and Jackson actually ran their experiment on at least five brands of carbon paper. Exhibit 8 summarizes, in square-unifying-table form, the results for all $\frac{1}{2}(5)(4) = 10$ comparisons among the five brands, and proceeds to find fitted values and combining residuals.

In this example we find that:

Midspread of the fitted differences = 49.2

Midspread of combining residuals = 10.2

so that, in relative terms, the quality of the fit for all

exhibit 7 of chapter 20 -- peanut proteins

Biological Value of Proteins in Raw and Roasted Peanuts as Calculated From Rat Growth

A) ANALYSIS

Pair of rats	Biological value: Raw	Roasted	Difference	Ordered Differences
1	61	55	6	-18
2	60	54	6	2
3	56	47	9	2H
4	63	59	4	2
5	56	51	5	3m
6	63	61	2	4m
7	59	57	2	5
8	56	54	2	6H
9	44	62	-18	8
10	61	58	3	9
				(Difference of raw & roasted)
Mean	57.9	55.8	2.1	(2.1 = diff of means)
Trimean	59.0	56.2	3.8	2.8 = diff of trimeans
(Median)	(59h)	(56.0)	(3h)	(3h = diff of median)

B) SOURCE: H. H. Mitchell, Wise Burroughs, Jesse R. Beadles 1936. The significance and accuracy of biological values of proteins computed from nitrogen metabolism data. Journal of Nutrition 11: 257-274. Table 1 on page 261.

exhibit 8 of chapter 20 -- 5 carbon papers

Analysis of the full carbon-paper experiment

Second brand	First brand A	B	C	D	E

A) DATA -- sums of 30 ratings, first brand over second brand

B	13				
C	-36	-21			
D	41	20	61		
E	2	-13	13	-40	

B) SQUARE UNIFYING TABLE

A	0	-13	+36	-41	-2
B	13	0	+21	-20	+13
C	-36	-21	0	-61	-13
D	41	20	61	0	+40
E	2	-13	13	-40	0
(Sum)	(20)	(-27)	(131)	(-162)	(38)
Mean	+4.0	-5.4	+26.2	-32.4	+7.6

C) FITTED DIFFERENCES

A	0	-9.4	+22.2	-36.4	3.6
B	9.4	0	31.6	-27.0	13.0
C	-22.2	-31.6	0	-58.6	-18.6
D	36.4	27.0	58.6	0	40.0
E	-3.6	-13.0	18.6	-40.0	0

D) COMBINING RESIDUALS

A	0	-3.6	13.8	-4.6	-5.6
B	3.6	0	-10.6	7.0	0.0
C	-13.8	10.6	0	-2.4	5.6
D	4.6	-7.0	2.4	0	0.0
E	5.6	0.0	-5.6	0.0	0
(median)	(3.6)	(0.0)	(0)	(0.0)	(0.0)

Clarity of combination = 11^+ (see later exhibit)

five carbon papers is not as good as for the three we looked at first, the latter midspread being 1/5 that of the former. (Notice that we should include the two combining residuals shown as ".0" in calculating the midspread of the combining residuals. They do not have to vanish, the way the five shown as "0" have to.)

The use of + signs in panel B is worth noting. In copying in the direct values from panel A, we continued to write, for example, "41" rather than "+41". This is surely clear enough, and in this copying situation, safe enough. In filling in the upper right portion of panel B, however, where the entries are the negatives of the symmetrically placed entries in the lower left portion, we have always used the "+".

Experience shows that we are likely to make too many errors of sign for comfort when filling out this part of the table. Insisting on a "+" or "-" for each entry reduces the number of errors we make. This is quite well worthwhile.

* polished combination *

Now that we have more than three versions to compare, we can go another step forward if we wish.

When we look at the combining residuals -- every column of which must sum to zero because of the way we have done the arithmetic -- we have an impression of unsymmetrical

straggling. In order to take this further, the last line of exhibit 8 contains the medians of the columns of residuals. There continues to be a suggestion that we might be wise to consider taking account of this appearance.

These residuals offer a convenient way to do just that. The original fit to brand A was +4.0. The median combining residual was +3.6. Accordingly it would be natural to try a modified fit of

$$+7.6 = +4.0 + 3.6$$

for brand A, and corresponding values for the other brands. Exhibit 9 carries this out, producing new fits and new residuals. This time the bulk of the combining residuals are substantially smaller, and we are inclined to trust the fitted values further, especially for the definition of residuals.

Clearly, we could repeat this process of median polishing of the combination as many times as we wish. (There would be no change in this example.) It is probably safest, as a routine procedure, to do it

only once.

Exceptional situations may require more, but every such instance should be examined from a number of points of view -- and thought through carefully. (Experience may lead to relaxing this position of caution.)

exhibit 9 of chapter 20 -- 5 carbon papers

Polished combination of the data of exhibit 8

	First brand				
	A	B	C	D	E
A) CALCULATION OF NEW FIT					
Means from A of 9	+4.0	-5.4	+26.2	-32.4	+7.6
Midmeans from C of 9	+3.6	0	0	0	0
Sums	+7.6	-5.4	+26.2	-32.4	+7.6
B) FITTED DIFFERENCES -- based on new fit					
A	0	-13.0	18.6	-40.0	+0.0
B	13.0	0	31.6	-27.0	+13.0
C	-18.6	-31.6	0	-58.6	-18.6
D	40.0	27.0	58.6	0	40.0
E	-0.0	-13.0	18.6	-40.0	0
C) COMBINING RESIDUALS -- see panel B of exhibit 8					
A	0	0.0	17.4	-1.0	-2.0
B	.0	0	-10.6	7.0	.0
C	-17.4	10.6	0	-2.4	5.6
D	1.0	-7.0	2.4	0	.0
E	2.0	0.0	-5.6	0.0	0.0
(median)	(0.0)	(0.0	(0)	(0.0)	(0.0)

D) CLARITY of COMBINATION

	Original analysis	Polished combination
Range of fitted differences	117.2	117.2
Midspread of combining residuals	10.2	4.4
Clarity of combination	11^+	26^+

* comparing the combining residuals *

We believe the combining residuals of panel C of exhibit 9 are better than those of panel C of exhibit 8. Let us look at them. Exhibit 10 displays their values, omitting all those that must be 0 because they are on the diagonal. As is so likely to happen, a dot plot tends to focus our attention on the outer parts of the batches, mainly reminding us that the very largest residuals are larger for the fuller treatment. This is not -- either here or frequently -- a satisfactory way to look at our residuals.

So that we may be clearer about what is going on, exhibit 11 shows the schematic plots for the same two sets of residuals. The increase in size of the largest combining residuals is still clear, but even clearer is the large shrinkage of the midspread, which goes from 10.2 to 4.4. If midspread of combining residuals is a useful criterion of the closeness of our incomplete description, as we tend to believe, the second set of combining residuals is very much better than the first.

Another way to see the comparison in even more detail is given by exhibit 12, in which the dots of the dot plots are connected by tracks showing their changes due to median polishing.

exhibit 10 of chapter 20 -- 5 carbon papers

The two sets of combining residuals in a dot plot
(diagonal zeros omitted)

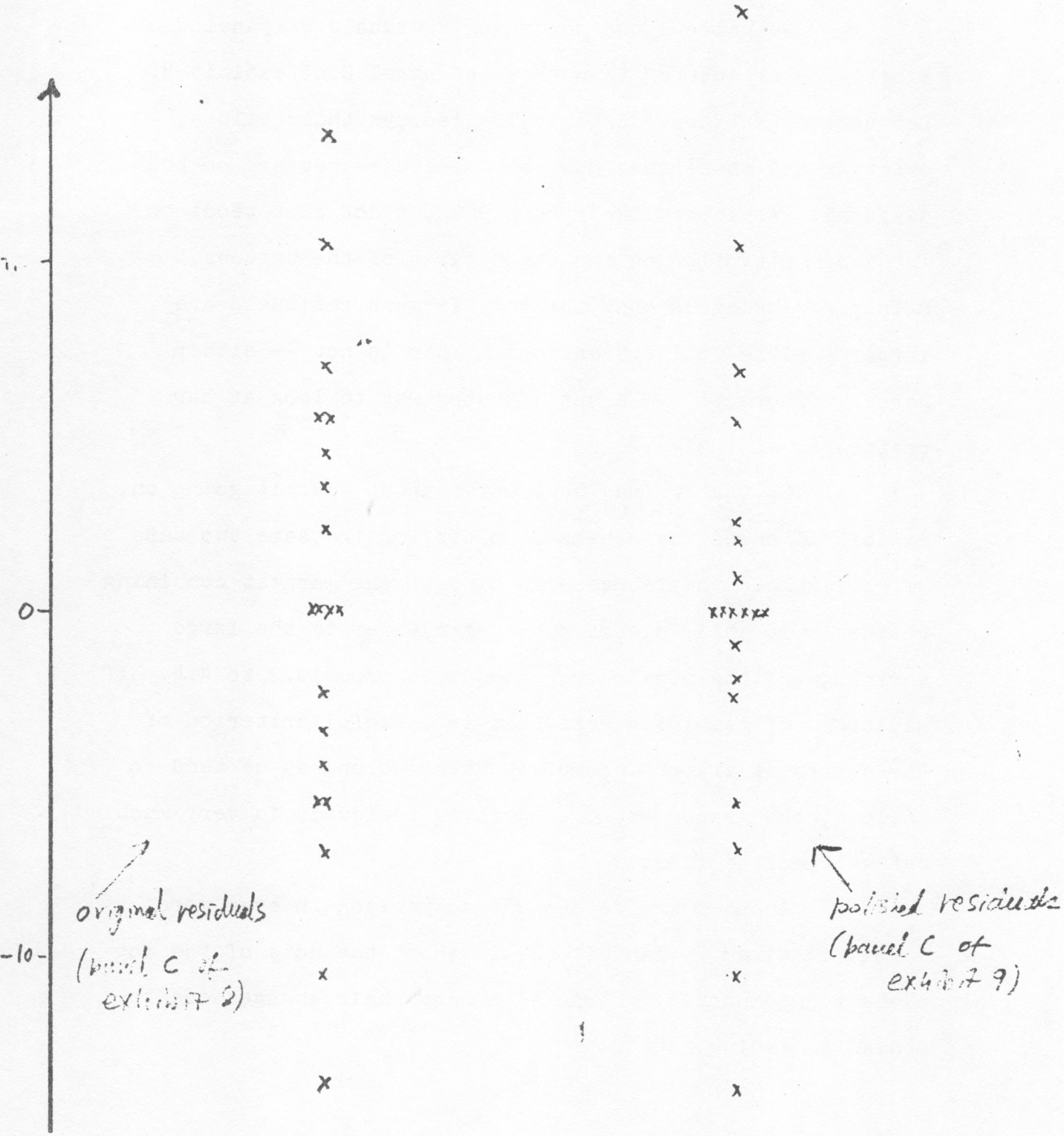


exhibit 11 of chapter 20 -- 5 carbon papers

Schematic plots for the two sets of residuals

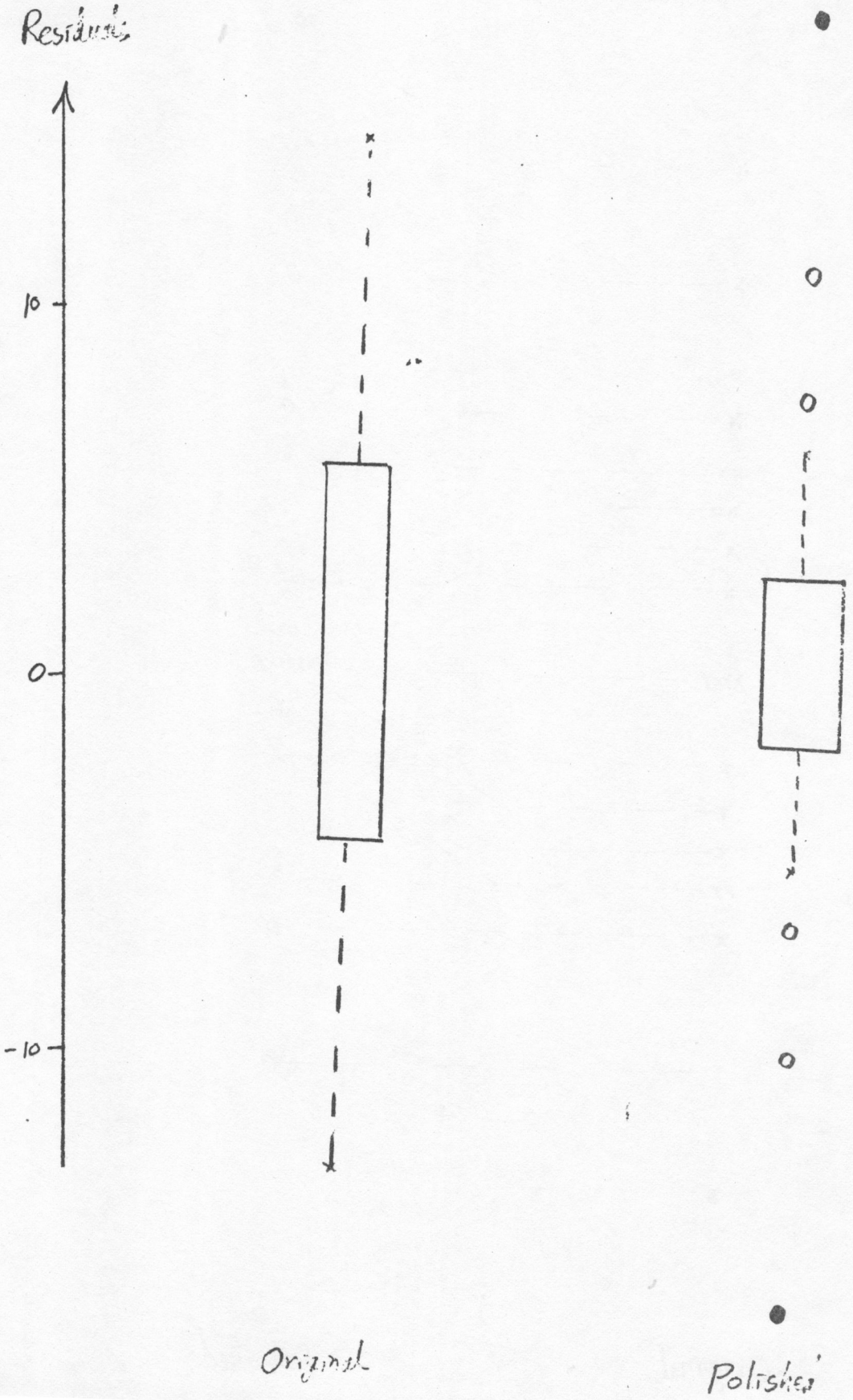
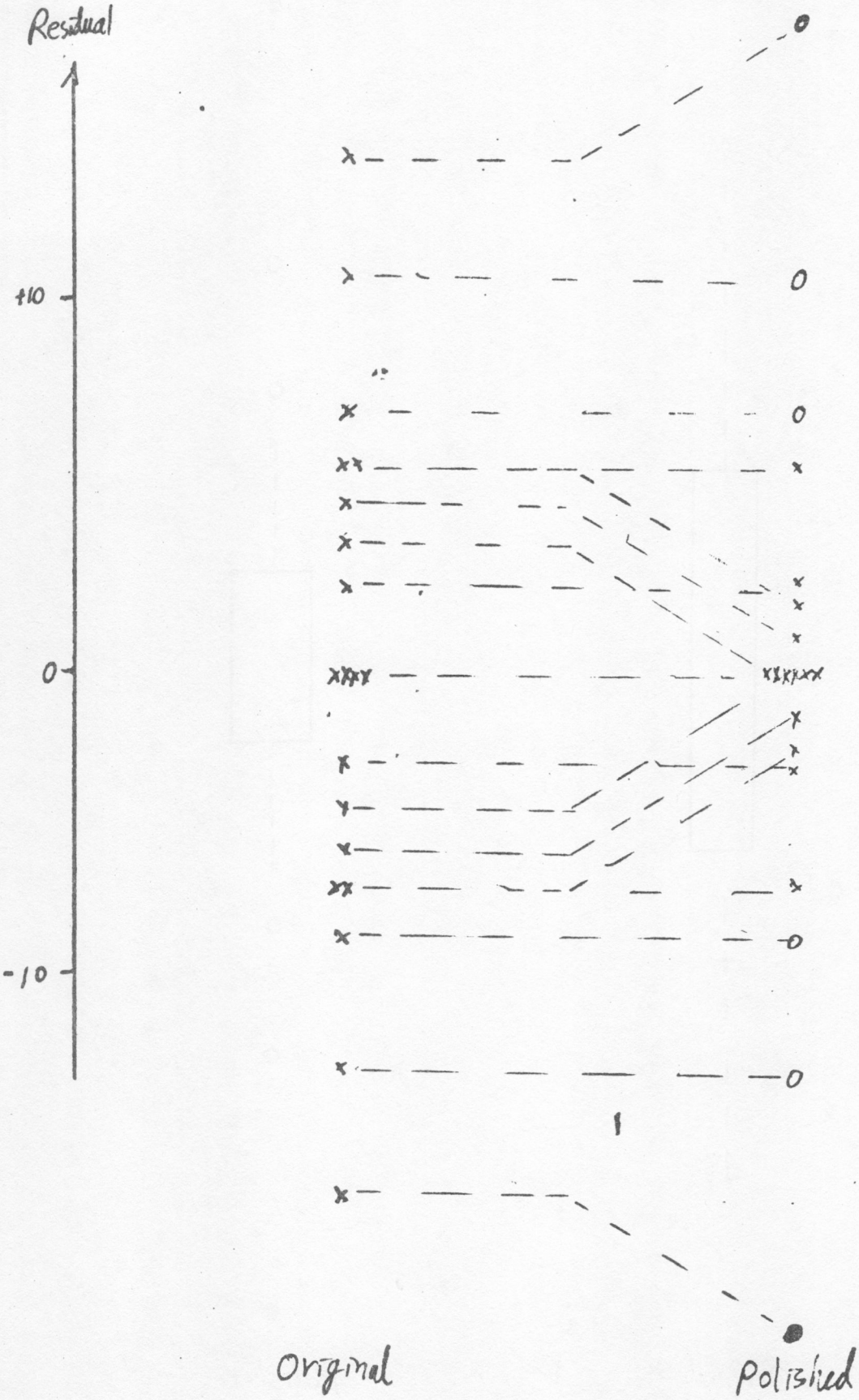


exhibit 12 of chapter 20 -- 5 carbon papers

The two sets of combining residuals in dot plots
with changes shown
(compare exhibits 10 and 11)



20E. <u>More than two, but less than all</u>

We have looked at two extremes:

- only comparisons in pairs.
- all candidates compared simultaneously.

Obviously there can be an intermediate situation where each planned comparison involves more than two but fewer than all of the candidates. Clearly we can apply to that situation the process we have been evoloving step by step, namely:

1) Look at all occurrences of any one comparison, note the corresponding differences, form the medians.

2) Enter these medians -- and their negatives -- in a square unifying table, find fitted differences, combining residuals, and median combining residuals.

3) If the median combining residuals seem large enough to concern us, add them to the original fits to form a new set of fits, and complete the process of fuller treatment by finding new fitted differences and new combining residuals.

Exhibit 13 shows a part of data reported by Eden and Fisher on the yield of straw and grain under each of eight manurial treatments. In this illustration, the incompleteness of comparison, which could have been part of the design, has been enforced by our selection of data. The problem of analysis is, of course, not altered by the way in

exhibit 13 of chapter 20 -- 5/8 straw yields

Yields of straw under eight different manurial treatments

A) DATA

Block	Treatment							
	A	B	C	D	E	F	G	H
1	242	-	-	-	322	200	260	203
2	321	382	-	-	-	261	318	275
3	261	201	298	-	-	-	266	207
4	317	316	381	255	-	-	-	331
5	255	280	300	238	232	-	-	-
6	-	285	294	309	393	258	-	-
7	-	-	256	283	351	306	276	-
8	-	-	-	324	363	376	385	328

B) DIFFERENCES, and their MEDIANS

A-B	A-C	A-D	A-E	A-F	A-G	A-H
-61	-37	62	-80	42	-18	39
60	-64	17	23	60	+3	46
1	-45				-5	54
-25						-14
-12	-45	40	-28	51	-5	42

B-C	B-D	B-E	B-F	B-G	B-H	C-D
-97	61	48	121	64	107	126
-65	42	-108	27	-65	-6	62
-20	-24				-15	-15
-9						-27
-42	42	-30	74	0	-6	24

C-E	C-F	C-G	C-H	D-E	D-F	D-G
68	36	32	91	6	51	7
-99	-50	-20	50	-84	-23	-61
-95				-68	-52	
				-39		
-95	-7	6	70	-54	-23	-27

D-H	E-F	E-G	E-H	F-G	F-H	G-H
-76	122	62	119	-60	-3	57
-4	135	75	35	-57	-14	43
	45	-22		30	48	59
	-13			-9		57
-40	84	62	77	-33	-3	57

exhibit 13 (cont'd.)

Minus treatment	Plus treatment A	B	C	D	E	F	G	H
C) SQUARE UNIFYING TABLE								
A	0	+12	45	-40	28	-51	5	-42
B	-12	0	42	-42	30	-74	0	6
C	-45	-42	0	-24	95	7	-6	-70
D	40	42	24	0	54	23	27	40
E	-28	-30	95	-54	0	-84	-62	-77
F	51	74	-7	-23	84	0	33	3
G	-5	0	6	-27	62	-33	0	-57
H	42	-6	70	-40	77	-3	57	0
Sum	43	50	85	-250	430	-215	54	-209
Mean	5	6	11	-31	54	-27	7	-26
D) FITTED DIFFERENCES								
A	0	+1	+6	-36	+49	-32	+2	-31
B	-1	0	+5	-37	+48	-33	+1	-32
C	-6	-5	0	-42	43	-38	-4	-37
D	36	37	42	0	85	+4	38	+5
E	-49	-48	-43	-85	0	-81	-47	-80
F	32	33	38	-4	81	0	34	+1
G	-2	-1	4	-38	47	-34	0	-33
H	31	32	37	-5	80	-1	33	0
E) FIRST COMBINING RESIDUALS								
A	0	11	39	-4	-21	-19	3	-11
B	-11	0	37	-5	-18	-41	-1	38
C	-39	-37	0	18	52	45	-2	-33
D	4	5	-18	0	-31	19	-11	35
E	21	18	-52	31	0	-3	-15	3
F	19	41	-45	-19	3	0	-1	2
G	-3	1	2	11	15	1	0	-24
H	11	-38	33	-35	-3	-2	24	0
Median	2	3	1	-2	-2	-1	-1	1
(Mean from C)	(5)	(6)	(11)	(-31)	(54)	(-27)	(7)	(-26)
Sum*	7	9	12	-33	52	-28	6	-25

* Results of polished combination.

F) SOURCE: T. Eden and R. A. Fisher 1927. Studies in Crop Variation. IV. The experimental determination of the value of top dressings with cereals. Journal of Agricultural Science 17: 548-562.

which the incompleteness comes about. Panels B, C, and D of exhibit 13 carry the analysis through the collection of differences, the formation of the square unifying table, and the calculation of fitted differences. As we collect differences, it is clear that some comparisons appear four times, some three, and some only twice. While we are likely to want to make some allowance for this in due course, the essential feature of the square unifying table -- differences of means used to get means of differences -- would make it hard for us to take account of these differences in weight at that point. Accordingly, we delay any account of weights.

Panel E presents the (first) combining residuals and, as a basis for fuller treatment, the medians of columns of residuals. The resulting fuller-treatment fits are not enough different from the first stage values to make us insist on a second set of combining residuals at this point. (We shall find it worthwhile to do this in a moment.)

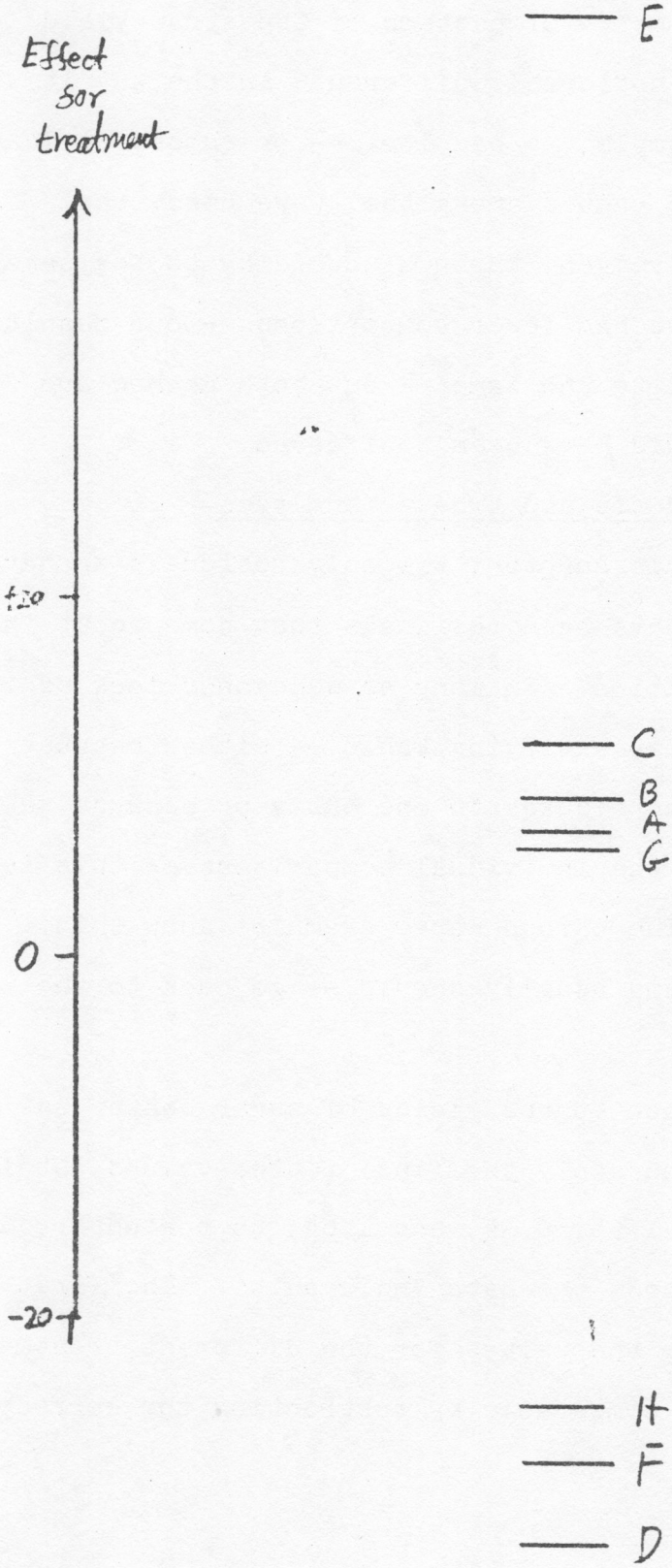
Exhibit 14 shows the fitted values visually. A reasonable verbal summary of the visual appearance would be:

- high yield of straw: E
- intermediate yield of straw: D,G,A,B
- low yield of straw: F,D,H

Bear in mind that this is just a summary of appearances. It has not been subjected to confirmatory analysis.

exhibit 14 of chapter 20 -- 5/8 straw yields

Relative amounts of straw by treatment
(as fitted in exhibit 13, at the close of panel E)



Before passing on to the next topic, we should emphasize that the symmetry of pattern of the straw yield data analyzed made no noticeable difference in the way it was analyzed. If, for example, we had lost -- or chosen to leave out -- almost any five values among the 40 we used, the whole process would have gone through according to the same pattern. We would have had fewer comparisons -- our results would not have been quite the same -- but both method and its reasonableness would have been unaffected.

20F. Individual residuals and two-way analyses

So far in this chapter, the only residuals we have formed and looked at have been residuals that come about in the process of combination. So long as we cannot look behind the entries in the square unifying table -- either because only one comparison contributes to one entry or because we do not have access to the individual comparisons -- this is all that we can do. The carbon paper examples show this. Otherwise, we can -- and usually should -- go back to the original comparisons.

Let us go back to the yields of sugar beets, as given in exhibit 4, and apply the final fitted values obtained in exhibit 6, namely -1.98, 0.34, and 1.64, to the observed values. Exhibit 15 shows the data and results. The first step is to try to make the values for the different treatments in a block reasonably comparable by subtracting the corresponding

fitted values. Once this far, it is easy and natural to find medians of these adjusted values, one for each block, and then subtract these medians to form individual residuals. As panel E shows, three of the residuals look rather "unusual" -- indeed, two are detached.

We have now made essentially all the calculations necessary for a moderately resistant analysis of this data as a two-way table. The fitted adjustments by treatment of panel B can serve as treatment effects, while the medians by block of panel C can serve as block fits. We need only to calculate their median, given as 5.18 in panel C, and then complete the elementary analysis as in exhibit 16. While the block effects are small, the overall pattern of the analysis is just what we have become familiar with as the pattern of an elementary analysis.

We can do the same sort of thing with incomplete data. Exhibit 17 takes the analysis of the straw yields of exhibit 12 and carries it on to a full elementary analysis. More interestingly, exhibit 18 shows the data of the full carbon-paper experiment in two-way table form. To do this, we had to be a little arbitrary. The first result, +13 for A over B, has been interpreted as +10 for A and -3 for B. The others have been treated separately. So long as block effects and block fits are not of any real interest, this does no real harm. All block ab says, in exhibit 18, is that A was found 13 rating points better than B.

exhibit 15 of chapter 20 -- sugar beets

Calculation of individual residuals

Block	Treatment: 0, none	1, phosphate	2, phosphate and nitrate
A) OBSERVED VALUES -- see exhibit 4			
1	2.45	6.71	6.48
2	2.25	5.44	7.11
3	4.38	4.92	5.88
4	4.35	5.23	7.54
5	3.42	6.74	6.61
6	3.27	4.74	8.86
B) FITTED ADJUSTMENTS (exhibit 7, mean forced to 0)			
	-1.98	0.34	1.64
C) ADJUSTED VALUES -- observed minus adjustments			
			medians
1	4.43	6.37	4.84 (4.84)
2	4.23	5.10	5.47 (5.10)
3	6.36	4.58	4.24 (4.58)
4	6.33	4.89	5.90 (5.90)
5	5.40	6.40	4.97 (5.40)
6	5.25	4.40	7.22 (5.25)
(median)			5.18
D) INDIVIDUAL RESIDUALS -- adjusted values minus block medians			
1	-.43	1.53	.00
2	-.87	.00	.37
3	1.78	.00	-.34
4	.43	-1.01	.00
5	.00	1.00	-.43
6	.00	-.85	1.97

exhibit 15 (cont'd.)

E) RESIDUALS as a BATCH -- unit 0.01

```
4     H|HHH
5     4|3
6     3|7
      2|
      1|
      0|
6     Z|000000
     -0|
     -1|
     -2|
6    -3|4
5    -4|33
3     L|LL
```

6 + (6) + 6 = 18, ✓, hinges at 5

Hinges:	- 43	and + 43
Midspread:	86	
Sides:	-129	and 129
*Outside:	0	and 3
(count/20):	(1)	

*(skeptically H97, +178, +153)

exhibit 16 of chapter 20 -- sugar beets

The elementary analysis corresponding to exhibit 15

Block	Fertilizer treatment 0	1	2	eff	fit
1	-.43	1.53	.00	-.34	4.84
2	-.87	.00	.37	-.08	5.10
3	1.78	.00	-.34	-.60	4.58
4	.43	-1.01	.09	.72	5.90
5	.00	1.00	-.43	.22	5.40
6	.00	-.85	1.97	.07	5.25
eff	-1.98	0.34	1.64	5.18	0.00
fit	3.20	5.52	6.82	0.00	-5.18

exhibit 17 of chapter 20 -- 5/8ths straw yields

Taking the straw yields to individual residuals
and an elementary analysis
(see exhibit 13)

A) THE DATA

Block	Treatment							
	A	B	C	D	E	F	G	H
1	242	-	-	-	322	200	260	203
2	321	382	-	-	-	261	318	275
3	261	201	298	-	-	-	266	207
4	317	316	381	255	-	-	-	331
5	255	280	300	238	232	-	-	-
6	-	285	294	309	393	258	-	-
7	-	-	256	283	351	306	276	-
8	-	-	-	324	363	376	385	328

B) TREATMENT EFFECTS -- fitted in panel C of exhibit 13

5	6	11	-31	54	-27	7	-26

C) ADJUSTED VALUES -- A minus B

	A	B	C	D	E	F	G	H	medians
1	237	-	-	-	268	227	253	229	(237)
2	316	376	-	-	-	288	311	301	(311)
3	256	195	287	-	-	-	259	233	(256)
4	312	310	370	286	-	-	-	357	(312)
5	250	274	289	269	178	-	-	-	(269)
6	-	279	283	340	339	285	-	-	(285)
7	-	-	245	314	297	333	269	-	(297)
8	-	-	-	355	309	403	378	354	(355)

D) INDIVIDUAL RESIDUALS -- C minus its block medians

	A	B	C	D	E	F	G	H
1	0	-	-	-	31	-10	16	-8
2	5	65	-	-	-	-23	0	-10
3	0	-61	31	-	-	-	3	-23
4	0	-2	58	-26	-	-	-	45
5	-19	5	20	0	-91	-	-	-
6	-	-6	-2	55	54	0	-	-
7	-	-	-52	17	0	36	-28	-
8	-	-	-	0	-46	48	23	-1

exhibit 17 (Cont'd.) - 2

E) ELEMENTARY ANALYSIS

	A	B	C	D	E	F	G	H	eff	fit
1	0	-	-	-	31	-10	16	-8	-53	237
2	5	65	-	-	-	-23	0	-10	21	311
3	0	-61	31	-	-	-	3	-23	-34	256
4	0	-2	58	-26	-	-	-	45	22	312
5	-19	5	20	0	-91	-	-	-	-21	269
6	-	-6	-2	55	54	0	-	-	-5	285
7	-	-	-52	17	0	36	-28	-	7	297
8	-	-	-	0	-46	48	23	-1	65	355
eff	5	6	11	-31	54	-27	7	-26	290	000
fit	295	296	301	259	344	263	297	264	000	-290

exhibit 18 of chapter 20 -- 5 carbon papers

The data of exhibit 8 put into two-way table form
(somewhat arbitrarily)

A) The DATA

Block	Brand				
	A	B	C	D	E
ab	10	-3	a	a	a
ac	-30	a	6	a	a
ad	40	a	a	-1	a
ae	0	a	a	a	-2
bc	a	-20	1	a	a
bd	a	20	a	0	a
be	a	-10	a	a	3
cd	a	a	60	-1	a
ce	a	a	10	a	-3
de	a	a	a	-40	0

B) PROBLEMS

1) Using the brand values of panel B of exhibit 8, carry this through to an elementary analysis.

2) Using the polished combination values of panel A of exhibit 9, do the same.

3) Make a conversion of the observed differences between pairs into (nominal) values for separate brands different from that in panel A above, and carry through to an elementary analysis. Compare the results with those found from the assignment of panel A.

* more general uses *

As both exhibit 17 and exhibit 18 show, the techniques of this chapter apply to two-way tables with so many absent values as to make the use of our earlier methods certainly tedious and plausibly at least a little doubtful. Indeed, as optional section 20K will show those self-selected readers who study it, similar techniques can go even further, dealing with two-way tables even closer to complete absence of comparisons.

We have developed these techniques in a special case -- where one of the two factors is "blocks", whose purpose is to provide more uniform conditions and thus -- by keeping "other things" more nearly constant -- to provide more precise comparisons. Techniques are not to be fenced in, however. Once we have taken the trouble to learn them we ought to try them out under suitably diverse circumstances.

Why should we not apply

differences

medians of differences

square unifying tables

polished combination

and the like to other two-way tables, to two-way tables in which neither factor plays the role of blocks? -- And the answer comes back "why not?"

Exhibit 19 tackles the Arizona-temperature data in just this way, comparing each pair of places by forming a difference within each month, and then taking medians over months. By doing this we are comparing temperatures and pairs of places within the same month, just as we compared yields of straw for pairs of treatments within the blocks. Months are not dull and uninteresting, as blocks usually are. Indeed, we want to analyze months as well as places. In analyzing places, however, we can -- and now do -- use months like blocks.

If it were not that 66 pairs of months makes the calculation rather large, we could interchange months and places, using places like blocks as we analyze months.

Any two-way table can be approached in this way. We have learned a new approach to the analysis of any two-way table, complete or shot full of absent values. When we come to analyzing two-way tables by computer, we will want to take this approach very seriously indeed. For hand calculation, since it does tend to need more effort, we are likely to use it only for small tables.

exhibit 19 of chapter 20 -- Arizona temperatures

A two-way table without "blocks" partly analyzed in the style of this chapter

A) TAGS 1 = Flagstaff, 2 = Prescott, 3 = Phoenix, 4 = Yuma

Temperatures are mean monthly temperatures at airports.

B) The DATA

Month	Place 1	2	3	4
Jan	25.3	35.3	49.7	55.3
Feb	28.8	39.4	54.3	60.1
Mar	34.4	44.8	59.9	65.7
Apr	42.1	52.9	67.1	72.8
May	50.0	60.9	75.5	80.4
Jun	58.3	70.1	83.9	87.8
Jul	65.2	76.0	90.1	94.6
Aug	63.4	73.9	88.3	93.7
Sept	57.0	68.2	82.7	88.3
Oct	46.1	57.0	70.8	76.4
Nov	35.8	45.6	58.4	64.2
Dec	28.4	37.8	52.1	57.1

C) DIFFS

	1 - 2	2 - 3	1 - 3	3 - 4	2 - 4	1 - 4
Jan	-10.0	-14.4	-24.4	-5.6	-20.0	-30.0
Feb	-10.6	-14.9	-25.5	-5.8	-20.7	-31.3
Mar	-10.4	-15.1	-25.5	-5.8	-20.9	-31.3
Apr	-10.8	-14.2	-25.0	-5.7	-19.9	-30.7
May	-10.9	-14.6	-25.5	-4.9	-19.5	-30.4
Jun	-11.8	-13.8	-25.6	-3.9	-17.7	-29.5
Jul	-10.8	-14.1	-24.9	-4.5	-18.6	-29.4
Aug	-10.5	-14.4	-24.9	-5.4	-19.8	-30.3
Sept	-11.2	-14.5	-25.7	-5.6	-20.1	-31.3
Oct	-10.9	-13.8	-24.7	-5.6	-19.4	-30.3
Nov	-9.8	-12.8	-22.6	-5.8	-18.6	-28.4
Dec	-9.4	-14.3	-23.7	-5.0	-19.3	-28.7
medians	-10.8h	-14.3h	-24.9h	-5.6	-19.6h	-30.3h

(cont'd.)

exhibit 19 (cont'd.)

D) UNIFYING TABLE

	1	2	3	4
1	0	10.8h	24.9h	30.3h
2	-10.8h	0	14.3h	19.6h
3	-24.9h	-14.3h	0	5.6
4	-30.3h	-19.6h	-5.6	0
(sum)	(-66.1h)	(-23.1h)	(33.7)	(55.6)
mean	-16.5h	-5.8	8.4	13.9

E) SOURCE

Climatography of the United States No. 60.2 Climates of the States. Arizona. U. S. Weather Bureau, Washington D. C., September 1959.

* transparency and insulation *

If all the observations in the same block are increased by exactly the same amount, all differences will be unchanged, as will their medians for other central values) and the results of submitting these medians to the square unifying table. Thus the treatment effects fitted by the approach of this chapter are insulated from uniform changes by blocks.

As for blocks and treatments, so too for two-way tables of any kind. If we use the approach of this chapter, each set of stripe effects are insulated from uniform changes of stripes of the other kind. Exhibit 20 sets out some illustrative problems.

20G. <u>"Weighting" by polishing</u>

When we were discussing exhibit 13 itself, we pointed out that the entries in the square unifying table came from differing numbers of observations and might thus deserve differing weights. We resolutely put that thought aside, and went on the easy way, indicating that we might come back to "weighting".

Why did the question arise? Because the two-way table was incomplete. What would "weights" need to reflect? The presence of some values and the absence of others. Can we now do this easily? Yes. How?

exhibit 20 of chapter 20 -- problems

Some instances and principles associated with the approach of this chapter

A) PROBLEMS

1) Work through a small numerical example to show that row effects by median differences are insulated from uniform changes (at least two) by columns.

2) Would the same be true if we took trimeans of differences? Means of differences? Midmeans of differences?

3) Work through a small numerical example to show that row effects by median differences are transparent to uniform changes (at least two) by row.

4) Would the same be true if we took trimeans of differences? Means? Midmeans?

5) Can we interchange rows and columns in problems (1) to (4)? Why? Why not?

6) Prove, through careful use of either words or formulas, that the insulation of problem (1) always happens.

7) Do the same for the transparency of problem (3).

8) Suppose we make a first, outer, two-way analysis in any way. Suppose we then make an inner analysis by median differences, as in this chapter. How will the resulting confined analysis compare with the result of analyzing the original data by median differences, as in this chapter?

9) What happens in problem (8) when "median differences, as in this chapter" is replaced by "median polishing"? By "trimean differences"?

10) Suppose an outer analysis is by median differences, as in this chapter, and that an inner analysis of the resulting residuals is made in a way that is prescribed in detail, will the combined analysis have the same properties of insulation and transparency as the analysis by median differences itself?

If, for example, we apply median polishing to the table of individual residuals, we will have taken care of the "weighting" we noted earlier, for our median polishing would take account of exactly the data that is present -- giving no attention to what is missing. Exhibit 21 has the arithmetic. The changes in the elementary analysis deserve attention, but, as panels E and F show, can hardly be called striking. Two of the eight treatments and two of the eight blocks have their fits changed appreciably.

exhibit 21 of chapter 20 -- 5/8ths straw yields

Median polishing the residuals of exhibit 17

A) The INDIVIDUAL RESIDUALS

Block	Treatment							
	A	B	C	D	E	F	G	H
1	0	-	-	-	31	-10	16	-8
2	5	65	-	-	-	-23	0	-10
3	0	-61	31	-	-	-	3	-23
4	0	-2	58	-26	-	-	-	45
5	-19	5	20	0	-91	-	-	-
6	-	-6	-2	55	54	0	-	-
7	-	-	-52	17	0	36	-28	-
8	-	-	-	0	-46	48	23	-1
medians	0	-2	20	0	0	0	3	-8

B) HALF-POLISHED VALUES -- panel A minus treatment medians

									medians
1	0	-	-	-	31	-10	13	0	0
2	5	67	-	-	-	-23	-3	-2	-2
3	0	-59	11	-	-	-	0	-15	0
4	0	0	38	-26	-	-	-	-53	0
5	-19	7	0	0	-89	-	-	-	0
6	-	-4	-22	55	54	0	-	-	0
7	-	-	-72	17	0	36	-31	-	0
8	-	-	-	0	-46	48	20	7	7

C) TWICE HALF-POLISHED VALUES -- panel B minus its block medians

1	0	-	-	-	31	-10	13	0
2	7	69	-	-	-	-21	-1	0
3	0	-59	11	-	-	-	0	-15
4	0	0	38	-26	-	-	-	-53
5	-19	7	0	0	-89	-	-	-
6	-	-4	-22	45	54	0	-	-
7	-	-	-72	17	0	36	-31	-
8	-	-	-	-7	-53	41	13	0

exhibit 18 (Cont'd.) - 2

D) CORRESPONDING ELEMENTARY ANALYSIS

	A	B	C	D	E	F	G	H	eff	fit
1	0	-	-	-	31	-10	13	0	-53	237
2	7	69	-	-	-	-21	-1	0	19	311
3	0	-59	11	-	-	-	0	-15	-49	243
4	0	0	38	-26	-	-	-	53	22	312
5	-19	7	0	0	-91	-	-	-	-21	269
6	-	4	-22	45	54	0	-	-	-5	285
7	-	-	-72	17	0	36	-31	-	2	292
8	-	-	-	-7	-53	41	0	0	72	362
eff	5	4	31	-31	54	-27	10	-34	290	000
fit	295	294	321	259	344	263	300	256	000	-290

Examples: 5=5+0, 4=6+(-2), 31=11+20,...(row effects)
295=295+0, 294=296(-2), 321=301+20,...(row fits)
-53=-53+0, 21=21+0, -47=-34+(-13),...(col effects)
237=237+0, 311=311+0, 243=256+(-13),...(col fits)

E) CHANGES in TREATMENT FITS

	treatments							
	H	D	F	A	B	C	G	E
exhibit 17	264	259	263	295	296	301	297	344
exhibit 18	256	"	"	"	294	321	300	"
change	-8	0	0	0	-2	+20	+3	0

F) CHANGES in BLOCK FITS

	blocks							
	1	3	5	6	2	4	7	8
exhibit 17	237	256	269	285	311	312	297	355
exhibit 18	"	"	"	"	309	"	292	362
change	0	0	0	0	-2	0	-5	7

* * * * *

20H. How far have we come?

The central ideas of this chapter are four:

- comparisons of one version with another, and their combination.
- blocks, as a natural element of data taking.
- the need for making comparisons compatible, and the usefulness of unifying tables in doing this.
- even with incomplete comparisons in blocks, we can both find a residual for each individual observation and complete an elementary analysis.

Most data gathering is for comparison -- either there-and-then or with more distant results. Comparison within the given body of data is very common and very important. Good ways to do it are needed. Comparisons, in the first instance, are of one with another, yet comparison among several is something we do frequently. Convenience and compactness are served when we can assign a single number to each candidate or treatment such that the difference of two such numbers is a good comparison of the two corresponding versions -- candidates or treatments or what have you. Such a set of numbers -- when fitting well enough -- is the natural and desirable incomplete description of such candidates or treatments for the comparison of each with all others.

Comparisons are always under circumstances. These circumstances often change recognizably from one part of our

body of data to another. Precision of comparison is higher if the data yielding numbers to be differenced came from one set of circumstances. Comparisons are thus likely to be made within blocks. Blocks can offer us both increased precision of comparison and increased diversity of circumstance.

While it is good to have a single number for each candidate or treatment -- numbers whose differences tell us the comparisons -- we do not get such a set of numbers without effort. Even direct comparisons of all pairs among as few as three candidates or treatments are not going to be exactly compatible when separately measured.

A convenient and effective way to enforce compatibility is the square unifying table.

The unifying table gives us

combining residuals.

These are useful, but tell only a part of the story.

If we go back and adjust the original values for the fit, and then allow for blocks, we can calculate

individual residuals.

They tell much more of the story. (They do not do as well as the combining residuals in telling what the combining residuals say.) Together with the quantities fitted in finding them, they provide the basis for an elementary analysis.

Polishing by medians can be applied either to the combining residuals, where we usually half-polish once, or to the individual residuals, where we usually half-polish twice, first by treatments and then by blocks.

The resistant analysis which this chapter provides for special two-way tables where one factor is "blocks" is an unsymmetric analysis in which:

- treatments are fitted by first looking at comparisons two-by-two, then using a unifying table to get the fitted treatment effects.
- after adjusting the data for these treatment effects, we find block fits as middle means by blocks.
- subtracting these fits leave individual residuals.
- once a common term is found as a mean -- or median -- of the block fits, it is easy to complete an elementary analysis.

20K. Where the unifying table has to wait (optional)

If the available comparisons are too few in kind, we cannot apply the square unifying table to our direct comparisons. Instead we have to combine direct and indirect comparisons slightly more tediously. And then, since the results of these combinations are NOT automatically compatible, we need to use the square unifying table to enforce compatibility.

Exhibit 22 takes the remainder of the Eden and Fisher data and carries out the process, continuing with median polish to provide "weighting".

Clearly there is some interest in comparing the results found here for 3/8 of the original data with those found in section 20F for the other 5/8. Exhibit 23 makes the comparison. The agreement, as we might expect, is far from perfect. It is good enough, however, to suggest that confirmatory data analysis would cause us to give a small but definite amount of credence to the results. In particular, treatment D seems to produce less straw than the others, while treatments E and G seem to have produced more straw. (There is a hint that median polish has brought the results closer together, but only a hint.)

We need to say again that we have not claimed -- and would not claim -- that the analysis set out here is "the best" or "the most precise".

At our stage of experience, we should be happy to have one way of analyzing such data that is moderately reasonable.

exhibit 22 of chapter 20 -- 3/8ths straw yields

The other part of the Eden and Fisher data and its analysis

A) The DATA

Block	Manurial treatment							
	A	B	C	D	E	F	G	H
1	-	267	215	212	-	-	-	-
2	-	-	330	292	370	-	-	-
3	-	-	-	265	284	259	-	-
4	-	-	-	-	323	361	340	-
5	-	-	-	-	-	234	362	229
6	331	-	-	-	-	-	400	266
7	216	200	-	-	-	-	-	276
8	295	295	284	-	-	-	-	-

B) The DIRECT COMPARISONS

A-B	A-C	A-G	A-H	B-C	B-D	B-H	C-D
16	11	-69	65	52	55	-76	3
0			-60	11			38
8	11	-69	2	32	55	-76	20

C-E	D-E	D-F	E-F	E-G	F-G	F-H	G-H
-40	-78	6	25	-17	21	5	133
	-19		-38		-128		134
-40	-98	6	-6	-17	-54	5	134

C) The DIRECT and INDIRECT COMPARISONS

A with B

direct	8	
	8	
via C	-21	= 11 - 32
via H	78	= 22- (-76)
median	8	

A with C

direct	11	
	11	
via B	40	= 8 - (-32)
	11	

A with D

via B	63	= 8 - (-55)
via C	31	= 11 - (-20)
median	47	

A with E

via C	-29	= 11 - (40)
via G	-52	= - 69 - (-17)
	-41	

(The other 24 comparisons go in the same ways, the results are summarized in panel D, following.)

exhibit 22 (Cont'd.) - 2

D) UNIFYING TABLE -- for the results of combined comparison illustrated in panel C.

Minus Treat	Plus treatment A	B	C	D	E	F	G	H
A	0	-8	-11	-47	+41	+9	+69	-2
B	8	0	-32	-55	+26	+10	+144	+76
C	11	32	0	-22	+40	+10	+68	+58
D	47	55	22	0	+98	-6	+82	+60
E	-41	-26	-40	-98	0	-16	+17	-58
F	-9	-10	-10	6	16	0	+54	-5
G	-69	-144	-68	-82	-17	-54	0	-102
H	2	-76	-58	-60	58	5	102	0
Sum	-46	-177	-199	-358	257	-42	536	29
Mean	-6	-22	-25	-45	32	-5	67	4

E) REMARK

Since we will plan to "weight" by midmean polishing, there seems to be no need for fuller treatment at this stage.

F) DATA ADJUSTED for TREATMENT -- A minus means in D

	A	B	C	D	E	F	G	H	median
1	-	289	240	257	-	-	-	-	257
2	-	-	355	337	338	-	-	-	338
3	-	-	-	310	252	264	-	-	264
4	-	-	-	-	291	366	273	-	273
5	-	-	-	-	-	239	295	225	239
6	337	-	-	-	-	-	333	262	333
7	222	222	-	-	-	-	-	272	222
8	302	317	309	-	-	-	-	-	309

exhibit 22 (Cont'd.) - 3

G) RESIDUALS

	A	B	C	D	E	F	G	H
1	-	32	-17	0	-	-	-	-
2	-	-	17	-1	0	-	-	-
3	-	-	-	56	-12	0	-	-
4	-	-	-	-	18	93	0	-
5	-	-	-	-	-	0	56	-14
6	4	-	-	-	-	-	0	-71
7	0	0	-	-	-	-	-	50
8	-7	12	0	-	-	-	-	-
(median)	0	12	0	0	0	0	0	-14

H) FIRST ELEMENTARY ANALYSIS

Block	Fertilizer treatment								eff	fit
	A	B	C	D	E	F	G	H		
1	-	32	-17	0	-	-	-	-	-22	257
2	-	-	17	-1	0	-	-	-	59	338
3	-	-	-	56	-12	0	-	-	-15	264
4	-	-	-	-	18	93	0	-	-6	273
5	-	-	-	-	-	0	56	-14	-40	239
6	4	-	-	-	-	-	0	-71	54	333
7	0	0	-	-	-	-	-	50	-57	222
8	-7	12	0	-	-	-	-	-	30	309
eff	-6	-22	-25	-45	32	-5	67	4	279	000
fit	273	257	254	234	309	274	346	283	000	-279

I) HALF-POLISH of panel G

									(median)
1	-	20	-17	0	-	-	-	-	0
2	-	-	17	-1	0	-	-	-	0
3	-	-	-	56	-12	0	-	-	0
4	-	-	-	-	18	93	0	-	18
5	-	-	-	-	-	0	56	0	0
6	4	-	-	-	-	-	0	-55	0
7	0	-12	-	-	-	-	-	64	0
8	-7	0	0	-	-	-	-	-	0

exhibit 22 (Cont'd.) - 4

J) A further HALF-POLISH of panel I

	A	B	C	D	E	F	G	H
1	-	20	-17	0	-	-	-	-
2	-	-	17	-1	0	-	-	-
3	-	-	-	56	-12	0	-	-
4	-	-	-	-	0	75	-18	-
5	-	-	-	-	-	0	56	0
6	4	-	-	-	-	-	0	-55
7	0	-12	-	-	-	-	-	64
8	-7	0	0	-	-	-	-	-

K) POLISHED ELEMENTARY ANALYSIS corresponding to panel J

									eff	fit
1	-	20	-17	0	-	-	-	-	-22	257
2	-	-	17	-1	0	-	-	-	59	338
3	-	-	-	56	-12	0	-	-	-15	264
4	-	-	-	-	0	75	-18	-	12	291
5	-	-	-	-	-	0	56	0	-40	239
6	4	-	-	-	-	-	0	-55	54	333
7	0	-12	-	-	-	-	-	64	-57	222
8	-7	0	0	-	-	-	-	-	30	309
eff	-6	-10	-25	-45	32	-5	67	-10	279	000
fit	273	269	254	234	309	274	346	269	000	-279

L) RESIDUALS of panel J as a BATCH

1	-5	5
	-4	
	-3	
	-2	
5	-1	2728
7	-0	71
(10)	Z	0000000000
7	0	4
6	1	7
5	2	0
	3	
	4	
4	5	66
2	6	4
1	7	5

7 + 10 + 7 = 24, hinges at depth 6h

Hinges:	- 4	and +10h
Midspread:		14h
Sides:	-18h	and +25
Outside:	1	and 4
(count/20):		(1)

exhibit 23 of chapter 20 -- 3/8ths and 5/8ths straw yields

Comparison of Effects for Different Parts
of the Eden and Fisher Data

Piece	Pan/Exh	A	B	C	D	E	F	G	H	median
A) INITIALLY										
3/8	H of 20	-6	-22	-25	-45	32	-5	67	4	
5/8	{E of 16, C of 13}	5	6	11	-31	54	-27	7	-26	
wt'd mean		**1**	**-4**	**-2**	**-36**	**46**	**-22**	**30**	**-15**	
(size of diff)		(11)	(28)	(34)	(14)	(22)	(22)	(60)	(30)	25
B) AFTER SIMPLE MIDMEAN POLISH										
3/8	K of 2	-6	-10	-25	-45	32	-5	67	-10	
5/8	D of 18	5	4	31	-31	54	-27	10	-34	
wt'd mean		**1**	**-1**	**10**	**-36**	**46**	**-19**	**31**	**-26**	
(size of diff)		(11)	(14)	(56)	(16)	(22)	(22)	(57)	(24)	22

****

20P. Problems for chapter 20

1-10. Do the same-numbered problems of exhibit 20, page 20-50.

11. Tucker, Wakeley and Cochran (1951) determined the effect of washing and removing excess moisture by wiping or by air current on the ascorbic acid content of turnip greens. The data are given in exhibit 24. Analyze and comment.

12. Bing (1953) compared the effect of several herbicides on the weights of the spikes produced by flowering gladiolus plants. Exhibit 25 gives the data. Analyze and comment.

13. Sherwood (1962) studied the influence of type of soil on the development of strength in mixtures of soil and cement. Exhibit 26 gives some of his data. Analyse the seven-day and 28-day data separately. Discuss the comparison.

14. Youden (1951) reports on the comparison of seven thermometers in sets of three. His description deserves attention as an example of intelligent care. Exhibit 27 gives the results after correction for time between readings. Analyze and comment.

15. Seliger (1950) reported on a more effective way of comparing radioactive standards adopted by the National Bureau of Standards (namely using blocks). Exhibit 28 gives some of his data. Analyze and comment.

16. Gulliksen and Tukey (1958), in discussing the confirmatory analysis of data consisting of paired comparisons among seven specimens of handwriting (taken from the Ayres scale). Exhibit 29 summarizes the judgments made by two groups, each of 100 students, one at Chicago in the late 30's and one at Princeton in the late 40's. Analyze the two sets of data. Repeat this, omitting all entries of +80 or more and all entries of -80 or less. Contrast the results of the analysis.

17. Mosteller (1951) used as an example the relative game-winning of the teams of the American League in 1948. Exhibit 30 gives the data. Analyze and comment.

18. Oesting and Kaufman (1945) presented data on the performance of a rapid method for the measurement of fat in meat. Exhibit 31 gives some of the data. Analyze and comment.

19. Gore (1952) gives comparative exposure lives for 16 coatings exposed four to a panel, both in Florida and in an accelerated aging device. Exhibit 32 gives the data. Analyze this as best you can neglecting the differences in RELATIVE rates of aging in Florida and in X-1. Comment on your analyses and others that might be made.

20. Do problem 1 of exhibit 18.

21. Do problem 2 of exhibit 18.

22. Do problem 3 of exhibit 18.

exhibit 24 of chapter 20

Milligrams of Ascorbic Acid Per 100 Grams
Dry Weight of Turnip Greens

Block	Control	Washed and blotted dry	Washed and dried in air current
1	950	857	917
2	887	1189	1072
3	897	918	975
4	850	968	930
5	975	909	954

Source: H. P. Tucker, J. J. Wakeley, and F. D. Cochran 1951. Effect of washing and removing excess moisture by wiping or air current on the ascorbic acid content of turnip greens. Southern Cooperative Service Bulletin 10: 54-56.

See also: R. G. D. Steel and J. H. Torrie 1960. Principles and Procedures of Statistics, page 137.

exhibit 25 of chapter 20

Average Weight of Gladiolus Spikes in Ounces

Block	2.4-D TCA	Check	DN/CR	Sesin
1	2.05	1.25	1.95	1.75
2	1.56	1.73	1.00	1.93
3	1.68	1.82	1.83	1.70
4	1.69	1.31	1.81	1.59

Source: A. Bing 1954. Gladiolus control experiments, 1953. The Gladiolus 1954 (New England Gladiolus Society, Inc.)

See also: R. G. D. Steel and J. H. Torrie 1960. Principles and Procedures of Statistics, page 137.

exhibit 26 of chapter 20

Unconfined Compressive Strength (in Pounds Per Square Inch) of 10% Cement - 90% Soil Mixtures After 7 and 28 Days

Batch	After 7 days				After 28 days			
	I	II	III	IV	I	II	III	IV
A	490	520	590	810	830	1020	970	1500
B	380	460	400	620	760	870	680	1210
C	345	45	20	395	720	750	860	1330
D	295	75	250	550	660	770	670	1040
E	380	60	20	90	670	940	620	1000
F	350	220	430	540	640	800	845	930
G	410	80	175	70	860	870	770	1130
H	335	265	500	905	750	980	830	1270

I = Leighton Buzzard sand

II = Tunstall Common A

III = Tunstall Common B

IV = Whitchurch

Source: P. T. Sherwood 1962. Rapid method for detecting the presence of deleterious organic matter in soil-cement. Journal of Applied Chemistry 12: 279-288.

See also: R. L. Wine 1964. Statistics for Scientists and Engineers.

exhibit 27 of chapter 20

Adjusted (For Time of Reading) Temperatures
Read on Seven Thermometers

(Entries Are Thousandths of a Degree Above 30.000°C)
(Tags identify thermometers A to G)

Set	Tags and readings		
1	A 56	B 31	D 35
2	A 58	E 16	F 41
3	B 41	C 53	E 24
4	B 46	F 46	G 32
5	C 54	D 43	F 50
6	A 68	C 60	G 34
7	D 50	E 32	G 38

Source: W. J. Youden 1951. Statistical Methods for Chemists, John Wiley and Sons. Table 55 on page 104.

exhibit 28 of chapter 20

Net Counting Rates for Radioactive Standards in Counts Per Second

Block	Sample			
	A	B	C	D
1	26.46	27.58	29.15	29.54
2	27.00	28.03	29.61	29.52
3	26.48	27.82	29.31	29.53
4	26.51	27.90	29.13	29.15

Source: H. H. Seliger 1950. A new method of radioactive standard calibration. Journal of Research of the National Bureau of Standards, 45: 496.

See also: W. J. Youden 1951. Statistical Methods for Chemists, page 88.

exhibit 29 of chapter 20

Preferences Between Handwriting Specimens
(Entries Are % Favored Minus % Not Favored)

Minus specimen	Plus specimen								
	50a	50b	50c	70a	70b	70c	80a	80b	80c
A) CHICAGO DATA									
50a	0	4	34	90	98	96	98	94	88
50b	-4	0	20	70	90	92	96	96	90
50c	-34	-20	0	52	56	84	82	72	92
70a	-90	-70	-52	0	52	74	90	58	56
70b	-98	-90	-56	-52	0	48	60	4	42
70c	-96	-92	-84	-74	-48	0	18	-48	12
80a	-99	-96	-92	-90	-60	-18	0	-70	-38
80b	-97	-96	-72	-58	-4	+48	70	0	22
80c	-94	-90	-92	-56	-42	-12	38	-22	0
B) PRINCETON DATA									
50a	0	4	32	76	96	96	94	66	72
50b	-4	0	20	38	94	92	86	88	82
50c	-32	-20	0	40	74	88	84	68	86
70a	-76	-38	-40	0	56	68	82	40	66
70b	-96	-94	-64	-56	0	28	56	-26	22
70c	-96	-92	-88	-68	-28	0	42	-40	16
80a	-94	-86	-84	-82	-56	-42	0	-70	-24
80b	-68	-88	-68	-40	26	-40	70	0	40
80c	-72	-82	-86	-66	-22	-16	24	-40	0

Source: H. Gulliksen and J. W. Tukey 1958. Reliability for the law of comparative judgement. Psychometrika 23, 95-110, specifically table 1 on page 99.

exhibit 30 of chapter 20

Fraction of Games Won Between Individual American League Teams in 1948

Losing team	Winning Team							
	Clev.	Bost.	N. Y.	Phil.	Detr.	St. L.	Wash.	Chic.
Clev.	-	.478	.545	.273	.409	.364	.273	.273
Bost.	.522	-	.364	.455	.318	.318	.318	.364
N. Y.	.455	.636	-	.455	.409	.273	.227	.273
Phil.	.727	.545	.545	-	.545	.182	.364	.273
Detr.	.591	.682	.591	.455	-	.500	.273	.364
St. L.	.636	.682	.727	.818	.500	-	.545	.381
Wash.	.727	.682	.773	.636	.727	.455	-	.429
Chic.	.727	.636	.727	.727	.636	.619	.571	-

Source: F. Mosteller 1951. Remarks on the method of paired comparisons. III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. Psychometrika 16: 207-218.

exhibit 31 of chapter 20

% Fat in Meat by Standard and Rapid Methods

Sample	Meat	Standard method	Rapid method
1	Wieners	22.0	22.3
2	"	22.1	21.8
3	"	22.1	22.4
4	"	22.2	22.5
5	Chopped ham	24.6	24.9
6	Chopped pork	25.3	25.6
7	Chopped ham	25.3	25.8
8	"	25.6	26.2
9	"	25.6	26.1
10	"	25.9	26.7
11	Wieners	26.0	26.3
12	"	26.2	24.9
13	"	27.0	26.9
14	Chopped ham	27.3	28.4
15	"	27.7	27.1
16	Pork sausage links	41.5	41.4
17	"	41.6	41.4
18	Bulk pork sausage	45.5	45.5
19		48.5	48.2
20		49.1	47.5

Source: R. B. Oesting and I. P. Kaufman 1945. Industrial and Engineering Chemistry, Analytical Edition 17, 125.

See also: W. J. Youden 1947.

exhibit 32 of chapter 20

Aging results for 16 Surface Coatings
(Entries Give Coating Tag, Actual Life (in parenthesis),
Life as Fraction of Lives of Four Coatings on Panel)

Panel	Loc.				
1	Fla	A (4 mo) .091	B (10 mo) .227	C (14 mo) .318	D (16 mo) .364
2	X-1	A (150 hr) .133	B (275 hr) .245	C (350 hr) .311	D (350 hr) .311
3	Fla	E (18 mo) .214	F (18 mo) .214	G (21 mo) .250	H (27 mo) .322
4	X-1	E (325 hr) .109	F (510 hr) .171	G (650 hr) .218	H (1500 hr) .502
5	Fla	I (10 mo) .182	J (21 mo) .382	K (15 mo) .273	L (9 mo) .163
6	X-1	I (625 hr) .234	J (900 hr) .336	K (700 hr) .262	L (450 hr) .168
7	Fla	M (24 mo) .363	N (8 mo) .118	O (12 mo) .176	P (24 mo) .353
8	X-1	M (1200 hr) .390	N (300 hr) .098	O (625 hr) .203	P (950 hr) .309
9	Fla	A (3 mo) .055	E (15 mo) .278	I (15 mo) .278	M (21 mo) .389
10	X-1	B (450 hr) .177	F (600 hr) .235	J (1000 hr) .392	N (500 hr) .196
11	Fla	C (12 mo) .200	G (24 mo) .400	K (18 mo) .300	O (6 mo) .100
12	X-1	D (250 hr) .093	H (1200 hr) .449	L (400 hr) .150	P (825 hr) .308